SIAM J. NUMER. ANAL.
Vol. 4, No. 1, 1967
Printed in U.S.A.

THE CONJUGATE GRADIENT METHOD FOR LINEAR AND NONLINEAR OPERATOR EQUATIONS*

JAMES W. DANIEL†

Introduction. In the early 1950's, the conjugate direction and conjugate gradient (CG) methods were developed for the iterative solution of finite systems of linear algebraic equations [8], [11], [17], [18], [19]; later these methods were given a more general form [16] and partially extended to a Hilbert space setting [6]. Although the CG method was first applied to *nonlinear* equations in 1964 [4], little theoretical justification was given.

We present a study of the CG method for linear and nonlinear operator equations in Hilbert space. To conserve space the special theory of linear equations has in general been outlined only; the details may be found in [3].

After developing the general conjugate direction and gradient methods, we prove convergence of the CG method for linear equations and characterize this iteration as optimal in a certain sense, using this fact to derive improved rate of convergence estimates which compare very favorably with that for steepest descent. We outline further improvements in the convergence results and mention that the method can be used on more general operators, sketching an application to differential equations.

Because of purely notational complexities that tend to obscure the simple ideas involved in the general form of the CG method for nonlinear equations, we first present the theory in detail for the simplest special case and then sketch the modifications necessary to prove the theorems for the general case. The simple form of the method is shown to exhibit global convergence (which is not valid more generally); rate of convergence estimates are obtained near the solution. These results are extended to a larger class of iterative methods, including steepest descent; the "optimality" and the resulting high rate of convergence distinguish the CG method from the other methods in the class.

The numerical and practical problems, such as the accuracy to which c_n need be determined or the effect of making computationally convenient modifications in the iteration, will be analyzed in a later report.

1. Linear equations.

1.0. The method of conjugate directions. Unless explicitly stated otherwise, we assume that *M is a given bounded linear operator, with bounded*

* Received by the editors October 4, 1965, and in revised form June 27, 1966.

† Mathematics Research Center, United States Army, University of Wisconsin, Madison, Wisconsin. This research was partially supported by the National Science Foundation.

inverse, acting from $\mathfrak{H}$ *onto* $\mathfrak{H}$, *a separable real Hilbert space with inner product* $\langle \cdot , \cdot \rangle$. *We choose* H, K *to be positive definite self-adjoint linear operators from* $\mathfrak{H}$ *onto* $\mathfrak{H}$ *and define* $N \equiv M^*HM$; N clearly has the same characteristics as H, K. For any k in $\mathfrak{H}$ there exists a unique solution h to $Mx = k$; we attempt to solve this equation by minimizing an error function.

DEFINITION 1.0.1. $E(x) \equiv \langle r, Hr \rangle$, where $r = k - Mx$.

We note that $E(x) = \langle h - x, N(h - x) \rangle$; thus $E(x) \geqq 0$, $E(x) = 0$ if and only if $x = h$. Following the outline of [2] we can prove [3]:

PROPOSITION 1.0.1. *Let* $\{p_n\}_0^\infty$ *be a sequence of linearly independent elements of* $\mathfrak{H}$ *such that* $\langle p_i , Np_j \rangle = 0$ *if* $i \neq j$. *Let* x_0 *be arbitrary and*

$$x_{n+1} = x_n + c_n p_n , \qquad c_n = \langle M^*Hr_n , p_n \rangle / \langle p_n , Np_n \rangle .$$

Let $\mathfrak{B}_n$ *be the space spanned by* $p_0 , \cdots , p_{n-1}$ *and* $\mathfrak{B}$ *be the closure of* $\bigcup_1^\infty \mathfrak{B}_n$. *Then* x_n *minimizes* $E(x)$ *on* $x_0 + \mathfrak{B}_n$ *and converges to a point* x' *in* $x_0 + \mathfrak{B}$ *with* $\langle M^*Hr', z \rangle = 0$ *for all* z *in* $\mathfrak{B}$, *where* $r' = k - Mx'$. *If* $\mathfrak{B} = \mathfrak{H}$, *then* $x' = h = M^{-1}k$; *otherwise* x' *may be distinct from* h.

A set of directions p_n as described above is called a set of N-conjugate (or conjugate) directions; such directions can be generated most generally by a scheme of Hestenes [16]:

PROPOSITION 1.0.2. *Let* K, N *be positive definite, bounded, self-adjoint linear operators in* $\mathfrak{H}$, *and let* $g_0 \neq 0$ *be in* $\mathfrak{H}$. *The algorithm*

$$(1.0.1) \qquad g_{n+1} = g_n - c_n N p_n , \qquad p_0 = K g_0 , \qquad p_{n+1} = K g_{n+1} + b_n p_n ,$$

where $c_n = \langle g_n , p_n \rangle / \langle p_n , Np_n \rangle$ *and* $b_n = -\langle Np_n , Kg_{n+1} \rangle / \langle p_n , Np_n \rangle$, *generates* N-*conjugate vectors* $g_0 , g_1 , \cdots$, *the sequence terminating at* $n = n_0$ *if and only if* $g_{n_0} = 0$. *Defining* $\mu(x) = \langle x, Nx \rangle / \langle x, K^{-1}x \rangle$, $\nu(x) = \langle x, Kx \rangle / \langle x, N^{-1}x \rangle$, $T = KN$, $a =$ inf spec (T), *and* $A =$ sup spec (T) *results in*

$$a \leqq \mu(p_i) \leqq \frac{1}{c_i} \leqq \mu(Kg_i) \leqq A,$$

and

$$a \leqq \nu(g_i) \leqq \frac{1}{c_i} \leqq \nu(Np_i) \leqq A.$$

Also $\langle g_{i+1} , p_i \rangle = 0$, $\langle g_i , p_i \rangle = \langle g_i , Kg_i \rangle$ *and* $b_i = \langle g_{i+1} , Kg_{i+1} \rangle / \langle g_i , Kg_i \rangle$.

1.1. The method of conjugate gradients. We consider the special conjugate direction method called the conjugate gradient method, derived by letting $g_0 = M^*Hr_0$ for an arbitrary x_0 and proceeding via (1.0.1). Clearly $g_n = M^*Hr_n$ for all n, so that (if $K = I$) p_n is generated by "conjugatizing" the negative of the gradient direction of $E(x)$ at $x = x_n$, i.e., g_n; thus the name.

By Proposition 1.0.1, x_n converges to x' minimizing $E(x)$ on $x_0 + \mathfrak{B}$, where $\mathfrak{B}$ is the closure of the span of $p_0, p_1, \cdots$; in the case of the CG method we guarantee that $x' = h$ as follows, using a trivial lemma.

LEMMA 1.1.1. $E(x_i) - E(x_{i+1}) = c_i\langle g_i, Kg_i\rangle = c_i^2 \langle p_i, Np_i\rangle$.

THEOREM 1.1.1. *$E(x_{i+1}) \leqq E(x_i)(1 - a/A)$. If K commutes with N, then $E(x_{i+1}) \leqq E(x_i)((A - a)/(A + a))^2$.*

Proof.

$$E(x_i) = \langle r_i, Hr_i\rangle = \langle M^{-1}r_i, M^*Hr_i\rangle = \langle M^{-1}(M^*H)^{-1}M^*Hr_i, M^*Hr_i\rangle$$

$$= \langle N^{-1}g_i, g_i\rangle = \frac{1}{\nu(g_i)}\langle g_i Kg_i\rangle.$$

$$E(x_i) - E(x_{i+1}) = c_i\langle g_i, Kg_i\rangle = E(x_i)c_i\,\nu(g_i).$$

The first estimate of the theorem follows from $c_i \geqq 1/A$, $\nu(g_i) \geqq a$. If K and N commute, then

$$a_i\,\nu(g_i) \geqq \frac{\nu(g_i)}{\mu(Kg_i)} = \frac{[Kg_i, Kg_i]^2}{([Kg_i, TKg_i][Kg_i, T^{-1}Kg_i])},$$

where $[x, y] \equiv \langle x, K^{-1}y\rangle$. It is easy to see that T is self-adjoint positive definite relative to $[\cdot\,,\,\cdot]$ with spectral bounds a, A; thus

$$c_i\,\nu(g_i) \geqq \frac{4aA}{(A + a)^2}$$

by the inequality of Kantorovich [10].

COROLLARY 1. *If α is the lower spectral bound for N, then*

$$\| x_n - h \|^2 \leqq \frac{E(x_0)}{\alpha}\left(1 - \frac{a}{A}\right)^n;$$

if K commutes with N, then

$$\| x_n - h \|^2 \leqq \frac{E(x_0)}{\alpha}\left(\frac{A - a}{A + a}\right)^{2n}.$$

Now we state our main theorem.

THEOREM 1.1.2. *The sequence $\{x_n\}$ generated by the CG method converges in norm to $h = M^{-1}k$.*

Proof. By Corollary 1, $\| x_n - h \|$ approaches zero.

1.2. Conjugate gradient method as an optimal process. E. Stiefel in [12], [20] used the theory of orthogonal polynomials to discuss "best" iterative methods for solving finite systems of linear algebraic equations; these results, extended to a Hilbert space setting, give another characterization of the CG method and lead to a much improved rate of convergence estimate.

Let M, H, N, K, a, A, E and T be as usual. We seek to solve $Mx = k$ by an iterative process. Practical considerations [3], [12] lead us to consider iterations of the form

$$x_{n+1} = x_0 + P_n(T)T(h - x_0), \tag{1.2.1}$$

where $P_n(\lambda)$ is a polynomial of degree n; to enable the use of spectral analysis for the rest of this section, we restrict ourselves to the case in which $N = \rho(T)$, where $\rho(\lambda)$ is a positive continuous function on the set $[a, A]$ containing spec (T). We seek $P_n(\lambda)$ for each n to minimize $E(x_{n+1})$ among all iterations of the form (1.2.1) given x_0. By the spectral theorem,

$$E(x_{n+1}) = \int_a^A \rho(\lambda)[1 - \lambda P_n(\lambda)]^2 \, ds(\lambda),$$

where $s(\lambda)$ is a known increasing function. The following generalization of a theorem of Stiefel follows in a straightforward fashion.

PROPOSITION 1.2.1. *The error measure $E(x_{n+1})$ is minimized by choosing $R_{n+1}(\lambda) \equiv 1 - \lambda P_n(\lambda)$ to be the* $(n + 1)$st *element of the orthogonal (on $[a, A]$, relative to $\lambda\rho(\lambda)ds(\lambda)$) set of polynomials $R_i(\lambda)$ satisfying $R_i(0) = 1$.*

If $K = I$ then Stiefel observed that a simple iteration exists to obtain this error; namely, if $R_{n+1}(q_i) = 0$, $i = 1, \cdots, n + 1$, then

$$x_{i+1} = x_i + \frac{1}{q_i} g_i = x_i + \frac{1}{q_i} M^*H(k - Mx_i)$$

yields the desired form for x_{n+1}. One can easily prove the following more general theorem by showing that the space $\mathfrak{V}_n$ of Proposition 1.0.1 is spanned by $T^0Kg_0, \cdots, T^{n-1}Kg_0$.

THEOREM 1.2.1. *The general CG method is the iteration of the form* (1.2.1) *minimizing $E(x_n)$ for all n among iterations of that form.*

Comparing the CG method to that of the form (1.2.1) derived by making $1 - \lambda p_n(\lambda)$ the $(n + 1)$st Chebyshev polynomial relative to $\lambda\rho(\lambda)ds$ on $[a, A]$ yields the following theorem.

THEOREM 1.2.2. *Let $\alpha = a/A$; then*

$$E(x_n) \leqq \left(\frac{2(1 - \alpha)^n}{(1 + \sqrt{\alpha})^{2n} + (1 - \sqrt{\alpha})^{2n}}\right)^2 E(x_0) \leqq 4\left(\frac{1 - \sqrt{\alpha}}{1 + \sqrt{\alpha}}\right)^{2n} E(x_0).$$

Clearly any other particular choice of $P_n(\lambda)$ will similarly yield a convergence estimate for the CG method; assuming only that spec $(T) \subset [a, A]$, we can do no better than Theorem 1.2.2. Under more specialized assumptions, it is possible to deduce more accurate estimates [9].

To solve $Mx = k$ by the method of steepest descent, one attempts to minimize $E(x) = \langle r, Hr \rangle$ proceeding from x_n to an x_{n+1} yielding the minimum of $E(x)$ in the direction of the negative gradient of $E(x)$ at $x = x_n$;

this is of the form (1.2.1), so we know that for each n, $E(x_n)$ is less for the x_n generated by the CG method with $K = I$ than for that by steepest descent. Since the best known [10], and in some cases best possible [1], estimate for the steepest descent method is

$$\| x_n - h \|^2 = \frac{1}{a} E(x_0) \left(\frac{A - a}{A + a}\right)^{2n},$$

while for the CG method we have now

$$\| x_n - h \|^2 = \frac{4}{a} E(x_0) \left(\frac{\sqrt{A} - \sqrt{a}}{\sqrt{A} + \sqrt{a}}\right)^{2n},$$

we see that the latter method appears to converge much more rapidly in general.

1.3. Special cases of the conjugate gradient method. The iteration is simplest when M itself is positive definite self-adjoint. Taking $H = M^{-1}$ and $K = I$ yields

$$N = T = M, \qquad E(x) = \langle h - x, M(h - x)\rangle.$$

Since $N = T$, we have $\rho(\lambda) \equiv \lambda$.

For general M, setting $H = K = I$ yields

$$T = N = M^*M, \qquad E(x) = \| r \|^2, \qquad \rho(\lambda) \equiv \lambda;$$

fortunately for computational uses one can avoid the computation of M^*M, [16].

A third special case which has been used arises from $H = (MM^*)^{-1}$, $K = M^*M$, yielding

$$N = I, \qquad T = M^*M, \qquad \rho(\lambda) \equiv 1, \qquad E(x) = \| h - x \|^2.$$

Again, by some manipulation, the iteration can be put in a convenient form [16].

In all three cases above, the form of $\rho(\lambda)$ assures the validity of the results of §1.2.

1.4. More rapid convergence; superlinear convergence. In the case of the first example of §1.3, R. Hayes [6] observed that for many operators improved rates of convergence can be obtained; this observation is also valid for the general method also, as Hestenes [16] has suggested.

Since $\langle g_i, Kg_j\rangle = 0$ if $i \neq j$, by Bessel's inequality we deduce that g_n' and Kg_n' converge weakly to zero in $\mathfrak{H}$, where $g_n' \equiv g_n/\langle g_n, Kg_n\rangle^{1/2}$. In the original rate of convergence proof (Theorem 1.1.1) the rate was given by an upper bound on $1 - \nu(g_n)/\mu(Kg_n)$. If K and N are perturbed by the addition of completely continuous operators, then, since g_n' and Kg_n'

converge weakly to zero, in the limit the ratios $\nu(g_n)$ and $\mu(Kg_n)$ are *not* perturbed. This produces the following theorem.

THEOREM 1.4.1. *Suppose $T = KN = P + C$, where C is completely continuous and P is a bounded linear operator such that K and $N' \equiv K^{-1}P$ are positive definite and self-adjoint, and let a_0, A_0 be the positive spectral bounds for P. Then $\| h - x_n \|^2$ and $E(x_n)$ for the CG method converge faster than any geometric series with ratio greater than*

$$1 - \frac{a_0}{A_0} = q.$$

If T and P are self-adjoint, then

$$q = \left(\frac{A_0 - a_0}{A_0 + a_0}\right)^2.$$

COROLLARY 1. *If there exists $\lambda > 0$ such that $T - \lambda I$ is completely continuous, then $E(x_n)$ and $\| h - x_n \|^2$ for the CG method converge faster than any geometric series with positive ratio.*

1.5. Extensions of the theory; applications. The above results can be easily extended to more general operators in several ways. For example, the results, slightly modified, hold [3], [6] for operators not everywhere definite, such as the Legendre operators of Hestenes [7]; however one must be careful in choosing the initial approximations in this case lest the iteraion at some point involve division by zero [3]. More importantly we can consider certain types of unbounded operators using the analysis developed by Friedrichs [5]. This well-known technique for "converting certain unbounded operators into bounded ones" has been clearly described in a general setting by Kantorovich [10].

As a simple example, suppose we seek to solve $Mx = k$, where

$$Mx = \sum_{i=1}^{N} \frac{\partial}{\partial s_i}\left(\alpha_i \frac{\partial x}{\partial s_i}\right) - \gamma x$$

is a linear elliptic differential operator on the set of functions x with Hölder continuous second derivatives in a bounded domain S and vanishing on the smooth boundary ∂S, where $\gamma = \gamma(s_1, \cdots, s_N)$ and the first derivatives of $\alpha_i = \alpha_i(s_1, \cdots, s_N)$ are Hölder continuous with $\alpha_i > 0$, $\gamma \geqq 0$ in $S + \partial S$. By using the boundedness and positive definiteness of M relative to the Dirichlet innerproduct, the problem can be reduced to one in which the earlier theory applies. The simplest CG iteration now takes the following form.

Pick x_0 with Hölder continuous second derivatives, $x_0 = 0$ on ∂S. Find $p_0 = r_0$ such that $\nabla^2 r_0 = Mx_0 - k$, $r_0 = 0$ on ∂S; let $d_0 = -\langle r_0, \nabla\rangle^2 r_0$

with the usual L^2 innerproduct over S. For $n = 0, 1, \cdots$, let $c_n = d_n/\langle p_n, Mp_n\rangle$, $x_{n+1} = x_n + c_n p_n$, find r_{n+1} such that $\nabla^2 r_{n+1} = \nabla^2 r_n + c_n M p_n$, $r_{n+1} = 0$ on ∂S, set $d_{n+1} = -\langle r_{n+1}, \nabla^2 r_{n+1}\rangle$, $b_n = d_{n+1}/d_n$, $p_{n+1} = r_{n+1} + b_n p_n$. Thus the solution of a general elliptic equation is reduced to the solution of a sequence of Poisson equations.

Further applications of the CG method, such as to matrix equations or integral equations, are more obvious and well known. We observe only that to solve the Fredholm integral equation $M\phi \equiv (I - \mu L)\phi = f$ for a μ less in magnitude than that of the eigenvalue of L with smallest magnitude, we obtain superlinear convergence according to the corollary to Theorem 1.4.1 since $M - I$ is completely continuous (for well-behaved kernels defining L).

2. Nonlinear equations.

2.0. The basic CG method; global convergence. Throughout this paper, unless otherwise stated, we assume that, in the standard terminology [13], *$J(x)$ is a continuous (in norm topology) operator in x, mapping the real Hilbert space $\mathfrak{H}$ into itself and having, at each point x in $\mathfrak{H}$, a bounded Frechet derivative J_x' with range $\mathfrak{H}$.* We seek to solve $J(x) = 0$.

By analogy with the linear case in which $J(x) = Mx - k$ and $J_x' = M$, we shall solve via directions conjugate relative to J_x' hoping to obtain the convergence found for linear equations at least near the solution. *Assuming* (through §2.2) *that J_x' is self-adjoint and satisfies $0 < aI \leqq J_x' \leqq AI$ for two real scalars a, A,* we shall consider the following iteration.

(2.0.1) Given x_0, let $p_0 = r_0 = -J(x_0)$; for $n = 0, 1, \cdots$ let $x_{n+1} = x_n + c_n p_n$, where c_n is the smallest positive root of $\langle J(x_n + c_n p_n), p_n\rangle = 0$; set $r_{n+1} = -J(x_{n+1})$, $p_{n+1} = r_{n+1} + b_n p_n$, where $b_n = -\langle r_{n+1}, J'_{n+1} p_n\rangle / \langle p_n, J'_{n+1} p_n\rangle$ and $J'_{n+1} \equiv J'_{x_{n+1}}$.

The assumptions on $J(x)$ and J_x' guarantee [13] that $J(x)$ is the gradient of a real valued function $f(x)$ defined and bounded below on $\mathfrak{H}$, also, for any x and p in $\mathfrak{H}$, a real c satisfying $\langle J(x + cp), p\rangle = 0$ must yield the minimum of $f(x + cp)$ as a function of c. The following lemma follows trivially from the assumptions on J_x' and from (2.0.1).

LEMMA 2.0.1. $\langle p_n, J_n' p_{n-1}\rangle = 0$; $\langle r_{n+1}, p_n\rangle = 0$; $\langle r_n, p_n\rangle = \| r_n \|^2$; $\langle p_n, J_n' r_n\rangle = \langle p_n, J_n' p_n\rangle$; $\| p_n \|^2 = \| r_n \|^2 + b_{n-1}^2 \| p_{n-1} \|^2$;

$$\langle r_n, J_n' r_n\rangle = \langle p_n, J_n' p_n\rangle + b_{n-1}^2 \langle p_{n-1}, J_n' p_{n-1}\rangle.$$

From the above follows trivially:

LEMMA 2.0.2. $\| r_n \|^2 \leqq \| p_n \|^2 \leqq (A/a) \| r_n \|^2$.

The bound $\| p_n \|^2 \leqq (A/a) \| r_n \|^2$ depends on the fact that p_n and p_{n-1}

are conjugate. Since we later wish to discuss iterations not using conjugacy, we will note the use of this assumption whenever it occurs.

THEOREM 2.0.1. *Under our standard assumptions, the sequence x_n generated by* (2.0.1) *is well defined for arbitrary x_0 and converges to the unique x' in $\mathfrak{H}$ satisfying $J(x') = 0$,*

$$f(x') = \min \{f(x);\ x \text{ in } \mathfrak{H}\}$$

and

$$\| x_n - x' \| \leqq \frac{1}{a} \| J(x_n) \|.$$

Proof. Let $f_n(c) = f(x_n + cp_n)$; then

$$f_n'(c) = \langle J(x_n + cp_n), p_n \rangle,$$

$$a \| p_n \|^2 \leqq f_n''(c) = \langle J'_{x_n+cp_n} p_n , p_n \rangle \leqq A \| p_n \|^2.$$

Since

$$f'(0) = -\langle r_n , p_n \rangle = - \| r_n \|^2 < 0,$$

we deduce that c_n exists and satisfies

$$c_n \geqq \frac{1}{A} \frac{\| r_n \|^2}{\| p_n \|^2} \geqq \frac{a}{A^2}.$$

Thus, for all $c \leqq c_n$, we have for some $0 \leqq t \leqq 1$,

$$\begin{aligned} f(x_n + cp_n) &= f(x_n) + c \langle J(x_n), p_n \rangle + \frac{c^2}{2} \langle J'_{x_n+tcp_n} p_n, p_n \rangle \\ &\leqq f(x_n) - c \| r_n \|^2 + \frac{1}{2} c^2 \frac{A^2}{a} \| r_n \|^2. \end{aligned}$$

Thus

$$\begin{aligned} f(x_{n+1}) &\leqq f\left(x_n + \frac{a}{A^2} p_n\right) \\ &\leqq f(x_n) - \frac{a}{A^2} \| r_n \|^2 + \frac{1}{2} \frac{a}{A^2} \| r_n \|^2 \leqq f(x_n) - \frac{1}{2} \frac{a}{A^2} \| r_n \|^2. \end{aligned}$$

Since $f(x)$ is bounded below, it follows that $J(x_n)$ converges to zero. Since

$$f(x) \geqq f(x_0) - \| x - x_0 \| \, \| J(x_0) \| + \tfrac{1}{2} a \| x - x_0 \|^2,$$

the set of x with $f(x) \leqq f(x_0)$ is bounded, hence $\| x_{n+k} - x_n \|$ is bounded; but

$$a \| x_{n+k} - x_n \|^2 \leqq \langle J(x_{n+k}) - J(x_n), x_{n+k} - x_n \rangle$$

which converges to zero. Thus there exists x' such that x_n converges to x'; clearly $J(x') = 0$ and $f(x') = \min \{f(x); x \text{ in } \mathfrak{H}\}$. Uniqueness follows from

$$\| J(x) - J(y) \| \| x - y \| \geqq \langle J(x) - J(y), x - y \rangle \geqq a \| x - y \|^2,$$

as does the speed estimate with $x = x'$, $y = x_n$.

2.1. Convergence near the solution. Hereafter we assume that *the assumptions of* §2.0 *are valid in a convex domain* $\mathfrak{B}$ *of* $\mathfrak{H}$ *in which we are always operating* ($\mathfrak{B}$ *will be defined later*), *and in addition that* J_x' *as a function of* x *has a linear Gateaux differential* J_x'' *satisfying* $\| J_x'' \| \leqq B$ *in* $\mathfrak{B}$. By $S(x, d)$ for positive d, x in $\mathfrak{H}$, we mean $\{y: \| x - y \| \leqq d\}$.

LEMMA 2.1.1. *If* $S(x_n, (1/a) \| r_n \|) \subset \mathfrak{B}$, *then* c_n *exists,*

$$\frac{a}{A^2} \leqq \frac{\| r_n \|^2}{A \| p_n \|^2} \leqq c_n \leqq \frac{1}{a} \frac{\| r_n \|}{\| p_n \|} \leqq \frac{1}{a},$$

and

$$\| c_n p_n \| \leqq \frac{1}{a} \| r_n \|.$$

Proof. $\langle J(x_n + cp_n), p_n \rangle = f_n'(c)$ of Theorem 2.0.1; by the argument of that theorem we deduce the result.

We notice that the lower bound a/A^2 comes from $\| p_n \|^2 \leqq (A/a) \| r_n \|^2$, which depends on the conjugacy relation.

LEMMA 2.1.2. $\| r_{n+1} \| \leqq (1 + A/a) \| r_n \|$ *if* $x_n + c_n p_n$ *is in* $\mathfrak{B}$.

Proof. $\| r_{n+1} \| \leqq \| r_{n+1} - r_n \| + \| r_n \| \leqq A \| x_{n+1} - x_n \| + \| r_n \| \leqq A \| c_n p_n \| + \| r_n \| \leqq (1 + A/a) \| r_n \|$.

DEFINITION 2.1.1. $E_n(x) \equiv \langle r, (J_n')^{-1} r \rangle$, $r = -J(x)$.

We note that

$$E_n(x) = \langle h_n - x_n, J_n'(h_n - x_n) \rangle,$$

where $h_n \equiv x_n + J_n'^{-1} r_n$ is the approximate solution given by Newton's method; thus our new error measure $E_n(x_n)$ measures our deviation from Newton's method. For convenience let $\epsilon_n^2 \equiv E_n(x_n)$. We now give four lemmas, the first of which is trivial, needed to prove the main theorem.

LEMMA 2.1.3. $(1/A) \| r_n \|^2 \leqq E_n(x_n) \leqq (1/a) \| r_n \|^2$.

LEMMA 2.1.4.

$$\frac{\| r_n \|^2}{\langle J_n' p_n, p_n \rangle} \cdot \frac{1}{1 + \eta_n} \leqq c_n \leqq \frac{\| r_n \|^2}{\langle J_n' p_n, p_n \rangle} \cdot \frac{1}{1 - \eta_n},$$

where $\eta_n = \frac{1}{2}(B\sqrt{A}/a^2)\epsilon_n$.

Proof. Let

$$g_n(c) = \langle J(x_n + cp_n), p_n \rangle = -\| r_n \|^2 + c\langle J_n' p_n, p_n \rangle + c^2 \int_0^1 (1 - t)\langle J''_{x_n + tcp_n}(p_n, p_n), p_n \rangle \, dt.$$

This gives

$$- \| r_n \|^2 + c\langle J_n' p_n , p_n\rangle - \tfrac{1}{2}c^2 B \| p_n \|^3 \leqq g_n(c)$$
$$\leqq - \| r_n \|^2 + c\langle J_n' p_n , p_n\rangle + \tfrac{1}{2}c^2 B \| p_n \|^3.$$

On the interval

$$0 \leqq c \leqq \frac{1}{a} \frac{\| r_n \|}{\| p_n \|},$$

this implies

$$- \| r_n \|^2 + c\langle J_n' p_n , p_n\rangle - \frac{1}{2} c \frac{1}{a} \frac{\| r_n \|}{\| p_n \|} B \| p_n \|^3 \leqq g_n(c),$$

and similarly above. Using $aI \leqq J_x'$ and $\| r_n \| \leqq \sqrt{A}\epsilon_n$ we deduce

$$- \| r_n \|^2 + c(1 - \eta_n)\langle J_n' p_n , p_n\rangle \leqq g_n(c),$$

and hence derive the upper bound on c_n. The lower bound is derived similarly.

LEMMA 2.1.5. *$E_{n+1}(x_{n+1}) - E_n(x_n) \leqq - c_n \| r_n \|^2 + d\epsilon_n^3$, where*

$$d = \frac{B}{a^3}\left(3 + \frac{A}{2a}\right).$$

Proof.

$$E_{n+1}(x_{n+1}) - E_n(x_n) = \langle r_n , (J_{n+1}'^{-1} - J_n'^{-1}) r_n\rangle + \langle r_{n+1} - r_n , J_{n+1}'^{-1} r_{n+1}\rangle$$
$$+ \langle r_{n+1} - r_n , J_{n+1}'^{-1} r_n\rangle \equiv X + Y + Z.$$

For the first term we have

$$\| J_{n+1}'^{-1} - J_n'^{-1} \| \leqq \frac{B}{a^2} \| c_n p_n \| \leqq \frac{B}{a^3} \| r_n \|,$$

so $| X | \leqq (B/a^3) \| r_n \|^3$. For

$$Y = \langle J(x_n) - J(x_{n+1}), J_{n+1}'^{-1} r_{n+1}\rangle$$
$$= \langle J_{n+1}'(x_n - x_{n+1}), J_{n+1}'^{-1} r_{n+1}\rangle + d_1 = d_1,$$

where, using an integral to represent d_1 as in Lemma 2.1.4 we have

$$| d_1 | \leqq \frac{1}{2} \frac{B}{a^3} \| r_n \|^2 \| r_{n+1} \|.$$

Using the same device we derive

$$Z = -c_n \| r_n \|^2 + d_2,$$

where

$$| d_2 | \leqq \frac{1}{2} \frac{B}{a^3} \| r_n \|^3.$$

Using $\| r_n \| \leqq \sqrt{A}\epsilon_n$ completes the proof.

LEMMA 2.1.6.

$$E_{n+1}(x_{n+1}) \leqq E_n(x_n)\left(\left(\frac{A-a}{A+a}\right)^2 + \sigma_n\right),$$

where

$$\sigma_n = \frac{4Aa}{(A+a)^2}\frac{\eta_n}{1+\eta_n} + d\epsilon_n .$$

Proof.

$$c_n \| r_n \|^2 \geqq \frac{\| r \|^2}{\langle J_n' p_n , p_n \rangle} \cdot \frac{1}{1+\eta_n} \cdot \frac{E_n(x_n)}{\langle r_n , J_n'^{-1} r_n \rangle} \cdot \| r_n \|^2.$$

Since $\langle J_n' p_n , p_n \rangle = \langle J_n' r_n , r_n \rangle + b_{n-1}^2 \langle J_n' p_{n-1} , p_{n-1} \rangle$, we find

$$c_n \| r_n \|^2 \geqq \frac{\langle r_n , r_n \rangle^2}{\langle J_n' r_n, r_n \rangle \langle J_n'^{-1} r_n, r_n \rangle} \cdot \frac{E_n(x_n)}{1+\eta_n} \geqq \frac{4aA}{(A+a)^2} \frac{E_n(x_n)}{1+\eta_n},$$

by the Kantorovich inequality [10]. Writing $E_{n+1}(x_{n+1}) = E_n(x_n) + (E_{n+1}(x_{n+1}) - E_n(x_n))$ and using Lemma 2.1.5 give the result.

We now prove our first theorem describing the local behavior of the convergence of the CG method.

THEOREM 2.1.1. *Assume* $0 < aI \leqq J_x' \leqq AI$, $\| J_x'' \| \leqq B$ *for all* x *in* $S(x_0 , R_0)$, *where* $R_0 = (\sqrt{A}/a(1 - q_0))\epsilon_0$ *and* x_0 *is chosen so that* $q_0^2 \equiv q^2 + \sigma_0 < 1$, $q = (A-a)/(A+a)$. *Then the CG iteration* (2.0.1) *generates a well-defined sequence* $\{x_n\}$ *converging to* x' *in* $S(x_0 , R_0)$ *with* $J(x') = 0$. $q_n^2 \equiv q^2 + \sigma_n$ *decreases to* q^2, *and*

$$\| x_n - x' \| \leqq R_0 q_0 q_1 \cdots q_{n-1} .$$

x' *is the unique solution in* $S(x_0 , R_0)$.

Proof. Our domain $\mathfrak{B}$ is $S(x_0 , R_0)$. For

$$| c | \leqq \frac{1}{a}\frac{\| r_0 \|}{\| p_0 \|}$$

we have $\| c p_0 \| < R_0$, so the assumptions above hold in a large enough domain about x_0 to make our lemmas valid for $n = 0$, and we can proceed to x_1. Let

$$R_1 = \frac{\sqrt{A}\epsilon_1}{a(1-q_1)} .$$

If x is in $S(x_1 , R_1)$, then

$$\| x - x_0 \| \leqq \| x - x_1 \| + \| x_1 - x_0 \|$$

$$\leqq \frac{\sqrt{A}\epsilon_1}{a(1-q_1)} + \frac{\sqrt{A}}{a}\epsilon_0 \leqq \frac{\sqrt{A}q_0\epsilon_0}{a(1-q_0)} + \frac{\sqrt{A}}{a}\epsilon_0 = R_0.$$

Thus $S(x_1, R_1) \subset S(x_0, R_0)$; our assumptions are now valid about x_1, and the constants η, σ, ϵ decrease. Continued induction gives

$$\frac{1}{A}\|J(x_n)\|^2 \leqq E_n(x_n) \leqq q_{n-1}^2 \cdots q_0^2\epsilon_0^2 \leqq q_0^{2n}\epsilon_0^2$$

and thus $J(x_n)$ converges to zero.

$$\|x_{n+k} - x_n\| \leqq \frac{1}{a}\sum_{i=n}^{n+k-1}\|r_i\| \leqq \frac{\sqrt{A}}{a}\epsilon_0 \sum_{i=n}^{n+k-1}(q_0 \cdots q_{n-1}),$$

proving that $\{x_n\}$ is a Cauchy sequence and yielding the existence of x'. Sending k to infinity gives

$$\|x' - x_n\| \leqq R_0 q_0 \cdots q_{n-1}.$$

Because of the optimality of the CG method for linear equations and the high rate of convergence guaranteed thereby, we expect that near the solution x' we should have more rapid convergence than that described in Theorem 2.1.1; this is indeed the case.

THEOREM 2.1.2. *Let the assumptions of Theorem 2.1.1 be satisfied. Then, for any m, there exists an integer n_m such that, for all $n > n_m$, we have*

$$E_{n+m}(x_{n+m}) \leqq (w_m^2 + \delta_n)E_n(x_n),$$

where

$$w_m = \frac{2\left(1 - \frac{a}{A}\right)^m}{\left(1 + \sqrt{\frac{a}{A}}\right)^{2m} + \left(1 - \sqrt{\frac{a}{A}}\right)^{2m}} \leqq 2\left(\frac{\sqrt{A} - \sqrt{a}}{\sqrt{A} + \sqrt{a}}\right)^m,$$

and δ_n tends to zero.

Proof. Consider the iterate x_n; let $z_0 = x_n$, z_i, $1 \leqq i \leqq m$, be the iterates derived by using the CG method to solve the linear equation $J_n'z = J_n'x_n + r_n$. Calling the solution $h_n = x_n + J_n'^{-1}r_n$, we have, by Theorem 1.2.2, that

$$\langle h_n - z_m, J_n'(h_n - z_m)\rangle \leqq w_m^2\langle r_n, J_n'^{-1}, r_n\rangle. \tag{2.1.1}$$

If we show that

$$|\langle h_n - z_m, J_n'(h_n - z_m)\rangle - E_{n+m}(x_{n+m})|,$$

which equals

$$|\langle h_n - z_m, J_n'(h_n - z_m)\rangle - \langle h_{n+m} - x_{n+m}, J'_{n+m}(h_{n+m} - x_{n+m})\rangle|,$$

is of order ϵ_n^3, then we will have

$$E_{n+m}(x_{n+m}) = \langle h_n - z_m , J_n'(h_n - z_m)\rangle$$
$$+ [E_{n+m}(x_{n+m}) - \langle h_n - z_m , J_n'(h_n - z_m)\rangle] \leqq (w_m^2 + O(\epsilon_n))E_n(x_n).$$

We indicate the proof of the order of magnitude. The sum to be estimated splits into

$$| \langle h_n - z_m , (J_n' - J'_{n+m})(h_n - z_m)\rangle |$$

and

$$| \langle h_n - h_{n+m} + x_{n+m} - z_m , J'_{n+m}(h_n - z_m + h_{n+m} - x_{n+m})\rangle | ,$$

the first of which is less than

$$B \| h_n - z_m \|^2 \| x_{n+m} - x_n \| = O(\epsilon_n^3),$$

by (2.1.1) and the proof of Theorem 2.1.1 (for big n). Clearly the second part of the sum is less than

$$\| h_n - h_{n+m} + x_{n+m} - z_m \| O(\epsilon_n);$$

we estimate the normed term. First

$$\| h_n - h_{n+m} \| = \| \sum_{i=0}^{m-1} (h_{n+i} - h_{n+i+1}) \| ,$$

while

$$\| h_{j+1} - h_j \| = \| x_{j+1} - x_j + J'^{-1}_{j+1} r_{j+1} - J'^{-1}_j r_j \|$$
$$= \| c_j p_j + J'^{-1}_{j+1}(r_{j+1} - r_j) + (J'^{-1}_{j+1} - J'^{-1}_j) r_j \| = O(\epsilon_n^2),$$

since

$$r_{j+1} - r_j = -J'_{j+1}(c_j p_j) + O(\epsilon_n^2).$$

We still must estimate

$$\| x_{n+m} - z_m \| = \| x_{n+m} + c_{n+m-1} p_{n+m-1} - z_{m-1} - \tilde{c}_{m-1} \tilde{p}_{m-1} \| ,$$

where the $\tilde{}$ indicates the z_i iteration. Since $\tilde{p}_0 = p_0$, a simple inductive argument yields

$$\| c_{n+i} p_{n+i} - \tilde{c}_i \tilde{p}_i \| = O(\epsilon_n^2)$$

for all i, which leads to

$$\| x_{n+m} - z_m \| = O(\epsilon_n^2).$$

2.2. A more general descent technique; steepest descent. In the previous sections we noted that the conjugacy relationships were only used to derive the inequality $\| p_n \|^2 \leqq (A/a) \| r_n \|^2$; of course Theorem 2.1.2 depends entirely on the conjugacy. If there exists C such that $\| p_n \|^2 \leqq C \| r_n \|^2$, then Theorems 2.0.1 and 2.1.1 are valid (with changes in the constants, convergence ratios, etc.). Since

$$\| p_n \|^2 = \| r_n \|^2 + b_{n-1}^2 \| p_{n-1} \|^2,$$

this condition can be replaced by

$$\| b_{n-1} p_{n-1} \| \leqq C \| r_n \| .$$

Thus we have the following theorems.

THEOREM 2.2.1. *Under the assumptions of Theorem* 2.0.1, *the iteration* (2.0.1), *modified so that* b_{n-1} *is chosen only so as to satisfy*

$$\| b_{n-1} p_{n-1} \| \leqq C \| r_n \| ,$$

generates a sequence $\{x_n\}$ *converging to the unique* x' *in* $\mathfrak{H}$ *with* $J(x') = 0$; $f(x') = \min \{f(x); x \text{ in } \mathfrak{H}\}$, *and* $\| x_n - x' \| \leqq \| J(x_n) \| /a$. *In particular this is valid for steepest descent, i.e.,* $b_{n-1} = 0$.

THEOREM 2.2.2. *Under the assumptions of Theorem* 2.1.1 *with* b_{n-1} *chosen as in Theorem* 2.2.1 *above,* x_n *converges to* x' *satisfying* $J(x') = 0$.

$$\| x_n - x' \| \leqq R_0 q_0 q_1 \cdots q_{n-1} \leqq R_0 q_0{}^n,$$

where q_i *decreases to*

$$q = \left(1 - \frac{1}{C} \frac{a}{A}\right)^{1/2}.$$

In particular this is valid for steepest descent, i.e., $b_{n-1} = 0$, *in which case we may derive* $q = (A-a)/(A+a)$.

The condition that $\| b_{n-1} p_{n-1} \| \leqq C \| r_n \|$ merely requires that the angle between r_n and p_n be bounded away from 90°; any such method will exhibit geometric convergence near the solution. The CG method, because of its "optimality" in the linear case, apparently has the smallest provable convergence ratio. Thus, in variational and minimization problems which demand numerical solution, it appears that the CG method will be a powerful tool of solution, certainly in many cases in which the steepest descent method is used.

2.3. The general CG method. Suppose that for each x, H_x and K_x are self-adjoint, uniformly (in x) bounded, uniformly positive definite linear operators; let

$$N_x = J_x'^* H_x J_x' , \qquad T_x = K_x N_x .$$

Suppose that $\| J_x'' \|$, $\| J_x' \|$, and $\| J_x'^{-1} \|$ are uniformly bounded and that K_x, N_x, T_x satisfy uniform Lipschitz conditions in x, i.e., $\| T_x - T_y \| \leqq d \| x - y \|$, etc.; we have also $0 < aI < T_x < AI$ for some real a, A. For convenience, H_n, T_n, etc. will denote H_{x_n}, T_{x_n}, etc.

Given x_0, let $r_0 = -J(x_0)$, $g_0 = J_0'^{*}H_0 r_0$, $p_0 = K_0 g_0$; for $n = 0, 1, \cdots$, pick c_n such that

$$\langle J(x_n + c_n p_n), H_n J_n' p_n \rangle = 0$$

(2.3.1) and let $x_{n+1} = x_n + c_n p_n$; $r_{n+1} = -J(x_{n+1})$; $g_{n+1} = J_{n+1}'^{*} H_{n+1} r_{n+1}$; set

$$p_{n+1} = K_{n+1} g_{n+1} + b_n p_n$$

where

$$b_n = -\langle K_{n+1} g_{n+1}, N_{n+1} p_n \rangle / \langle p_n, N_{n+1} p_n \rangle.$$

The iteration (2.3.1) defines the general CG method; we define error functions $E_n(x) \equiv \langle r, H_n r \rangle$. To prove convergence for this iteration we must prove the direct extensions of Lemmas 2.0.1, 2.0.2, 2.1.1–6; this is a trivial matter so we merely indicate the type of modifications necessary in the argument. For convenience let $\epsilon_n^2 = E_n(x_n)$. In Lemma 2.0.1 we have $\langle r_n, p_n \rangle = \| p_n \|^2$; the expected generalization would be $\langle g_n, p_n \rangle = \langle g_n, K_n g_n \rangle$. We deduce the slight perturbation

$$\begin{aligned}
\langle g_n, p_n \rangle &= \langle g_n, K_n g_n \rangle + b_{n-1} \langle g_n, p_{n-1} \rangle, \\
\langle g_n, p_{n-1} \rangle &= \langle J_n'^{*} H_n r_n, p_{n-1} \rangle = \langle (J_n'^{*} H_n - J_{n-1}'^{*} H_{n-1}) r_n, p_{n-1} \rangle \\
&= \epsilon_n \epsilon_{n-1} \| p_{n-1} \| O(1), \\
\| b_{n-1} p_{n-1} \| &= O(\epsilon_n),
\end{aligned}$$

so that

$$\langle g_n, p_n \rangle = \langle g_n, K_n g_n \rangle + O(\epsilon_n^2 \epsilon_{n-1}) = \langle g_n, K_n g_n \rangle (1 + O(\epsilon_{n-1})).$$

Proceeding in this fashion with all the lemmas we finally deduce that

$$E_{n+1}(x_{n+1}) \leqq q^2 E_n(x_n) + \sigma_n,$$

where $q = (A-a)/(A+a)$ if K_x and N_x commute, $q = (1 - a/A)^{1/2}$ otherwise, and where $\sigma_n = O(\epsilon_n^3 + \epsilon_n^2 \epsilon_{n-1} + \epsilon_n \epsilon_{n-1}^2 + \epsilon_{n-1}^3)$. From this it is simple to prove the following theorem.

THEOREM 2.3.1. *Let the above mentioned assumptions on K_x, H_x, N_x, T_x and $J(x)$ hold in a sphere of radius R_0 about the initial point x_0, where $R_0 = O(\epsilon_0)$ can be determined explicitly as in Theorem 2.1.1. If $E_0(x_0)$ is small enough, the sequence $\{x_n\}$ generated by (2.3.1) converges to x' with $J(x')$*

$= 0$. *We have*

$$\| x' - x_n \| \leqq R_0 q_0 \cdots q_{n-1} \leqq R_0 q_0^{\,n},$$

where q_n *decreases to* q *and* $q = (A-a)/(A+a)$ *if* $K_x N_x = N_x K_x$, $q = (1 - a/A)^{1/2}$ *otherwise.*

Arguing as in Theorem 2.1.2, we can also obtain the improved convergence estimates expected of the CG method.

THEOREM 2.3.2. *Under the assumptions of Theorem* 2.3.1 *and the additional restriction that* $N_x = \rho(T_x)$, *where* ρ *is positive and continuous on* $[a, A]$, *given* m *there exists* n_m *such that for* $N > n_m$,

$$E_{n+m}(x_{n+m}) \leqq (w_m^{\,2} + \delta_n) E_n(x_n)$$

for δ_n *tending to zero, and*

$$w_m = \frac{2\left(1 - \frac{a}{A}\right)^m}{\left(1 + \sqrt{\frac{a}{A}}\right)^{2m} + \left(1 - \sqrt{\frac{a}{A}}\right)^{2m}} \leqq 2\left(\frac{\sqrt{A} - \sqrt{a}}{\sqrt{A} + \sqrt{a}}\right)^m.$$

Besides the case already considered in §§2.0–2.2, in which $H_x = J_x'^{-1}$, $K_x = I$, the other examples of §1.3 apply here.

If we take $H_x = K_x = I$, then $E_n(x) = \| r \|^2$, and the CG method has an interesting relationship with least squares methods. If one defines $f(x) = \frac{1}{2} \| J(x) \|^2$, then

$$\langle f_x'' h, h \rangle = \langle J_x''(h, h), J(x) \rangle + \langle J_x'^* J_x' h, h \rangle,$$

which is approximately $\langle N_x h, h \rangle$ near the solution to $J(x) = 0$. Since it follows that asymptotically we are minimizing $f(x_n + c_n p_n)$ at each step, a good approach might be actually to determine c_n via this least squares approach; experiments are underway to see whether this allows us in practice to obtain convergence starting further from the solution.

As a final example let $H_x = K_x = J_x'^{-1}$; hence $T_x = I$ and $a = A = 1$. Since $p_n = J_n'^{-1} r_n$ this is a modification of Newton's method, which, by Theorem 2.3.1, exhibits superlinear convergence.

Acknowledgment. The author wishes to thank Professor M. M. Schiffer for his guidance.

REFERENCES

[1] H. AKAIKE, *On a successive transformation of the probability distribution and its application to the analysis of the optimum gradient method*, Ann. Inst. Statist. Math., 11 (1952), pp. 1–16.

[2] H. A. ANTOSIEWICZ AND W. C. RHEINBOLDT, *Conjugate direction methods and the method of steepest descent*, A Survey of Numerical Analysis, J. Todd, ed., McGraw-Hill, New York, 1962, pp. 501–512.

[3] J. W. DANIEL, *The conjugate gradient method for linear and nonlinear operator equations*, Doctoral thesis, Stanford University, Stanford, 1965.

[4] R. FLETCHER AND C. M. REEVES, *Function minimization by conjugate gradients*, Comput. J., 7 (1964), pp. 149–154.

[5] K. FRIEDRICHS, *Spektraltheorie halbbeschränkter Operatoren und Anwendung auf die Spektralzerlegung von Differentialoperatoren*, Math. Ann., 109 (1934), pp. 465–487.

[6] R. M. HAYES, *Iterative methods of solving linear problems in Hilbert space*, Nat. Bur. Standards Appl. Math. Ser., 39 (1954), pp. 71–104.

[7] M. R. HESTENES, *Applications of the theory of quadratic forms in Hilbert space to the calculus of variations*, Pacific J. Math., 1 (1950), pp. 525–582.

[8] M. HESTENES AND E. STIEFEL, *Method of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Standards Sect. B., 49 (1952), pp. 409–436.

[9] S. KANIEL, *Estimates for some computational techniques used in linear algebra*, to appear.

[10] L. V. KANTOROVICH, *Functional analysis and applied mathematics*, Uspehi Mat. Nauk., 3 (1948), pp. 89–185.

[11] C. LANCZOS, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bur. Standards Sect. B., 45 (1950), pp. 255–283.

[12] E. STIEFEL, *Relaxationsmethoden bester Strategie zur Lösung linearer Gleichungssysteme*, Comment. Math. Helv., 29 (1955), pp. 157–179.

[13] M. M. VAINBERG, *Variational Methods for the Study of Nonlinear Operators*, Holden-Day, San Francisco, 1964.

[14] M. Z. NASHED, *Iterative methods for the solutions of nonlinear operator equations in Hilbert space*, Doctoral thesis, University of Michigan, Ann Arbor, 1963.

[15] W. V. PETRYSHYN, *Direct and iterative methods for the solution of linear operator equations in Hilbert space*, Trans. Amer. Math. Soc., 105 (1962), pp. 136–175.

[16] M. R. HESTENES, *The conjugate gradient method for solving linear systems*, Proceedings of Symposia on Applied Mathematics, vol. VI, Numerical Analysis, McGraw-Hill, New York, 1956, pp. 83–102.

[17] C. LANCZOS, *Solution of systems of linear equations by minimized iterations*, J. Res. Nat. Bur. Standards Sect. B., 49 (1952), pp. 33–53.

[18] E. STIEFEL, *Über einige Methoden der Relaxationsrechnung*, Z. Angew. Math. Phys., 3 (1952), pp. 1–33.

[19] ———, *Recent developments in relaxation techniques*, Proceedings of the International Congress on Applied Mathematics, vol. I, North Holland, Amsterdam, 1954, pp. 384–391.

[20] ———, *Kernel polynomials in linear algebra and their numerical applications*, Nat. Bur. Standards Appl. Math. Ser., 49 (1958), pp. 1–22.

[21] M. Z. NASHED, *The convergence of the method of steepest descents for nonlinear equations with variational or quasi-variational operators*, J. Math. Mech., 13 (1964), pp. 765–794.

[22] ———, *On general iterative methods for the solutions of a class of nonlinear operator equations*, Math. Comp., 19 (1965), pp. 14–24.